
EFFECTS OF INDIVIDUAL DIFFERENCES ON THE SEMANTIC PROCESSING OF NONWORDS: NEW INSIGHTS FROM THE ENGLISH CROWDSOURCING PROJECT

¹Fabio Marson, ¹Rolando Bonandrini, ¹Simona Amenta, ¹Marco Marelli

¹Psychology Department, Milano-Bicocca University, Milan, Italy

fabio.marson@unimib.it

The literature on lexical processing have largely neglected nonwords, since such stimuli were typically considered meaningless. Conversely, recent studies suggest that nonwords carry traces of semantic information that affect their processing and that, in turn, attribution of meaning to out-of-vocabulary linguistic materials could represent a basic process in human language. Emergentist approaches suggest that such mechanism might be influenced by individual differences (ID) related to language experience (e.g., vocabulary size). However, IDs' role in affecting semantic activation elicited by nonwords has never been tested. Such research venue is now made possible by megastudies, such as the English Crowdsourcing Project (ECP): a non-speeded vocabulary test administered to ~580k participants, that collected behavioral performance for a large number of words and nonwords. In the present study we explored the extent to which individual experience with language affect semantic processing of nonwords, analyzing the responses to ~303k nonwords from ~300k participants obtained through the ECP.

From a computational point of view, semantic representations for nonwords can be extracted using Fasttext, a word-embedding technique that exploit sublexical information (namely, n-grams) to compute vector space models of semantic information. We computed orthography-to-semantic consistency (OSC), semantic neighborhood density (SND), and length of semantic vector (magnitude) for each nonword and used them to model individual reaction times (rts). Age, education, vocabulary size, and their interaction with these semantic metrics were included in the model to address the effect of ID on semantic sensitivity to unfamiliar stimuli.

Results showed that all semantic features affect rts, with OSC and magnitude having an interference effect while SND exerting a facilitatory effect. Importantly, all the semantic metrics of interest interacted with age, education, and vocabulary size, modulating the impact of nonword processing at the behavioral level. Results are consistent with the idea that the mechanisms underlying attribution of meanings to nonwords are sensitive to language exposure and vocabulary size.

Keywords: nonwords; semantics; individual differences; megastudy; language experience

¹ The work was supported by the European Union, ERC-COG-2022, BraveNewWord (grant number: 101087053)